FORMATION CLOUD COMPUTING

Présentation du cours

- Qu'est ce que le cloud ?
- Les types de cloud
- Sous le capot des clouds
- L'impact du cloud sur le business
- Cloud privé / cloud public
- L'architecture des applications dans le cloud
- La haute disponibilité dans le cloud
- Test, déploiement et exploitation
- Le futur du cloud computing

- Les 5 grands principe du NIST (National Institute of Standards and Technology):
 - On-demand self-service
 - Pas d'interaction humaine pour le déploiement d'un serveur (stockage, réseau, ..)

- Broad network access
 - Accès aux plates-formes via le réseau et via différents supports (Tablettes, smartphones, portables)

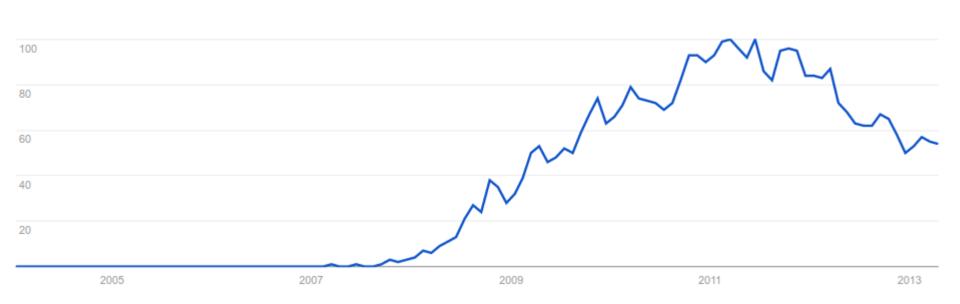
- Resource pooling
 - Les ressources (CPU, RAM, stockage, bande passante) sont mises en commun et permettent de fournir plusieurs clients (multi-tenant model). Les ressources peuvent « bouger » d'un endroit à un autre et sont virtuelles ou physiques.
- Rapid elasticity
 - Notion « d'illimité », les capacités peuvent augmenter ou diminuer (parfois automatiquement) dynamiquement
- Measured service
 - Transparence sur l'utilisation des ressources et facturation « pay-per-use »

- Une évolution de l'informatique :
 - 1960-1980 : Mainframe (et machines virtuelles VM-CP)
 - 1990 : Modèle client-serveur
 - 1996 : Première apparition du terme « cloud »
 - 2000 : Internet, début de la virtualisation (2001)
 - 2005+ : Début du cloud (Amazon, ..)
- Attentes importantes dans le secteur IT et en dehors

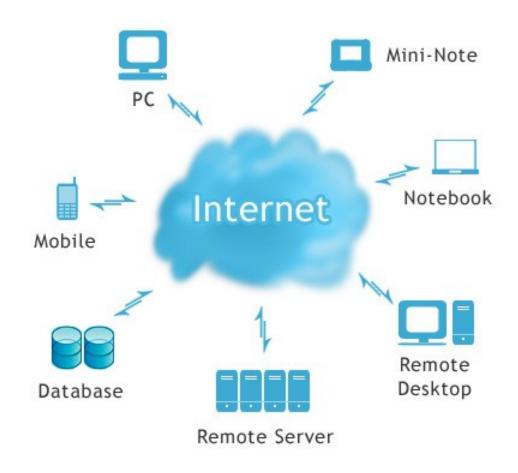
- Ce qui a permis au cloud d'exister :
 - Les technologies :
 - Virtualisation
 - Service-Oriented Architecture (SOA)
 - Augmentation des débits réseaux
 - **-** . . .

- Modèle économique :
 - SaaS

Evolution de l'intérêt des internautes (google)
 pour le terme « cloud computing » :



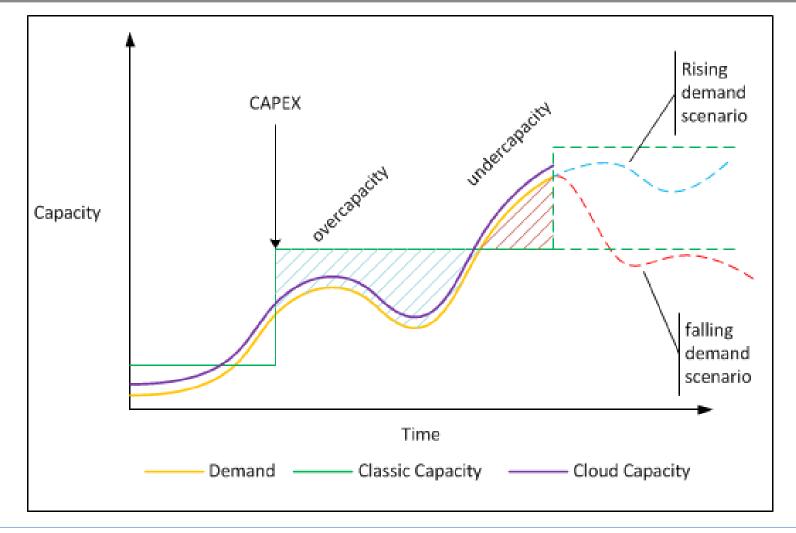
Pourquoi le « cloud »?



- Les bénéfices attendus :
 - Economies :
 - Ongoing operational costs (OPEX)VS
 - Up-front capital expenditure (CAPEX)
 - Agilité

- Les bénéfices attendus (suite) :
 - Meilleur « Time to market »
 - Moins d'investissements initiaux
 - Disponibilité immédiates des ressources

- Meilleur sécurité
 - Générateurs, accès internet multiple, bâtiments



Les types de clouds

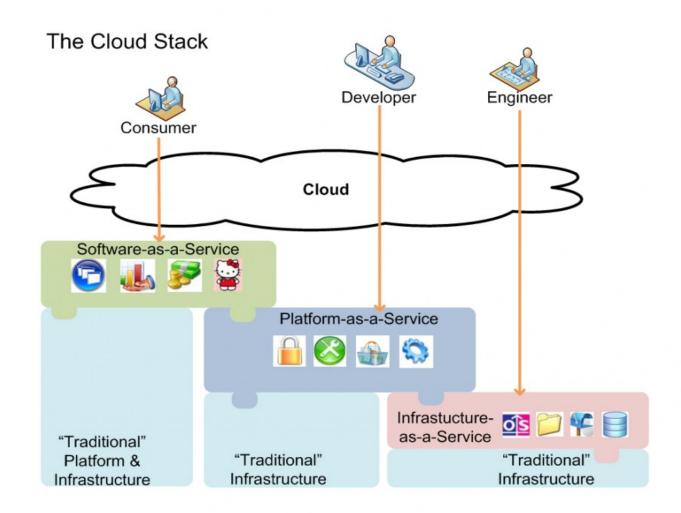
- Infrastructure as a Service (laaS ou HaaS)
 - Mise à disposition de machines virtuelles
 - Facturation à l'heure d'utilisation
 - Facture des volumes de données (stockage, réseau)
 - Exemple : Amazon, OVH, ...
- Platform as a Service (PaaS)
 - Mise à disposition de plates-formes (OS + application)
 - Facturation identique au mode laaS
 - Système « clé en main »
 - Exmeple Google AppEngine, Microsoft Azure, Heroku, ...

Les types de clouds

- Software as a Service
 - Mise à disposition d'un logiciel
 - Facturation au nombre d'utilisateur
 - Exemple : saleforces.com
- DCatacenter as a Service (DCaaS)
 - Location de salles dans un datacenter
 - Services ...

Desktop as a Service (DaaS)

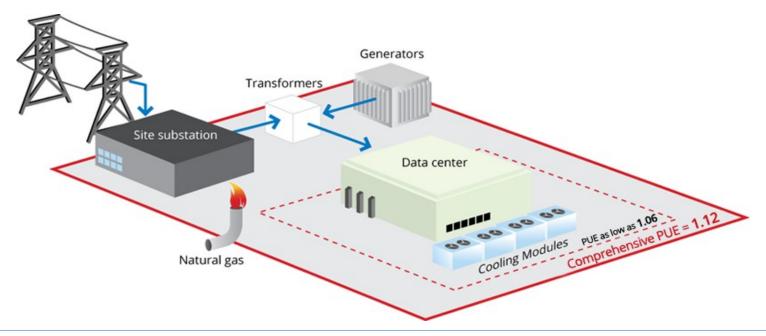
Les types de clouds



Sous le capot des clouds

Datacenter

- Tier 4
- Effet d'échelle
- PUE proche de 1 pour les meilleurs, 1,8-2,5 pour les autres



- Virtualisation
 - Serveurs, stockage, réseau
- Répartiteurs de charge
- API
 - Amazon, Azure
 - Representational State Transfer (REST)

Sous le capot

BDD

- XML, NoSQL (Cassandra, MongoDB, CouchDB)
- Fin de l'ACID (Atomicité, Cohérence, Isolation, Durabilité)

| Voiture | | | |
|---------|--|--|--|
| Clé | Attribut | | |
| 1 | Constructeur : Peugeot Modèle : 407 Couleur : Bleue Année : 2011 | | |
| 2 | Constructeur : Peugeot Modèle : 207 Couleur : Grise Année : 2010 Boîte : Automatique | | |

- Les modèles économiques
 - Hébergement interne
 - IT classique
 - Location d'espace
 - Plus de datacenter en interne, meilleure fiabilité
 - Service géré
 - Le sous traitant loue un matériel dédié au client et s'occupe de l'administration des équipements
 - Cloud
 - Equipements mutualisés, on ne paye que ce que l'on utilise

| Туре | Périmètre d'administration | Fiabilité | Coûts |
|------------------------|-------------------------------|-----------|-------|
| Hébergement interne | ++++ | +/++/+++ | ++ |
| Location d'espace | +++ | ++/+++ | ++ |
| Service géré | + | +/++/+++ | +++ |
| Cloud | ++ | ++/+++ | +/++ |

- Cas d'usages du cloud :
 - Besoins à court terme / limités dans le temps
 - Permet d'éviter de gros cout d'acquisition
 - Charge mouvante
 - Sites d'e-commerces, lancement de produits, ...
 - Applications non stratégiques
 - Petites applications web

- Le cloud n'a pas de sens pour :
 - Les vieilles plates-formes
 - ERP monolitique & co
 - Temps réel / « mission critical »
 - Problème avec l'approche best-efforts
 - Dépendance à Internet
 - Applications traitants de données confidentielles
 - Agréments de l'hébergeur
 - Localisation des données parfois impossible

Le cloud pour qui :

- Les startups aux faibles capitaux
 - N'ont pas les moyens de se payer des gros serveurs et n'en ont pas forcément besoin
- Les PME
 - Ne peuvent pas assurer un hébergement de serveurs aussi fiable et sécurisé que Google et manquant de compétences
- Les grandes entreprises
 - Peuvent profiter de l'agilité du cloud pour mettre en place des projets à forte capacité d'évolution tout en maitrisant les coûts

Cloud public

- Cloud public
 - Opérés par des spécialistes
 - Quel niveau de service ?
 - Quelles performances ?
 - Quel niveau de sécurité ?
 - Où sont les données ?
 - Patriot Act ?

- Les clouds publiques en action :
 - Amazon, Google, Microsoft, ...
 - Cloudwatt, Numergy

Cloud privé / cloud public

Cloud privés

- Opérés par les services informatiques internes
- Petits par nature
- Mise en commun des ressources, virtualisation, élasticité
- Sont 'ils vraiment des clouds ?
- Héritent des avancées du cloud public...
- …et sont forcément en retard
- Sont 'ils vraiment plus sécurisés ?
- Les clouds privés en actions :
 - CNRS, gouvernement américain, eBay, Facebook

La sécurité dans le cloud

- Sécurité physique
 - Datacenter avec de forts moyens

- Sécurité logique
 - Peut on faire mieux que Google en interne ?

- Les options de déploiement d'un cloud privé ;
 - Interne ou externe, VPC
 - Choix du produit
 - OpenSource
 - Eucalyptus, cloudstack, openstack, OpenNebula
 - Propriétaire
 - VMware, Enomaly
 - Taille (capacity planning)

- Architecture d'un cloud privé :
 - Portail utilisateur
 - Orchestrateur / contrôleur
 - Clusters
 - Nœuds
 - Stockage
 - Routage

- Choix des serveurs :
 - Approche « pod » ?





- Réseau
 - Simplification des zones
 - Utilisation des API des équipements
 - Isolation au plus proche des VM
 - Virtualisation réseau (ACI, NSX, ...

- Stockage
 - Utilisation de classes
 - Bronze, silver, gold, ..

- Les « patterns »
- Transference
 - Déplacer une application dans le cloud pour des raisons de coûts
- Internet scale
 - Volonté de créer une application pour le cloud pour un très grand nombre d'utilisateurs
- Burst compute
 - Capacité à déborder dans le cloud pour absorber les pics d'utilisation
- Elastic storage
 - Très gros besoins de stockage, difficilement prédictibles

- Le sharding
 - Découpage d'un base de données en plusieurs sous ensemble plus petits (des shards). Cette conception nécessite d'utiliser un modèle « shared-nothing ».

Pour les bases de données distribuées

 Pour résoudre des problèmes de latence propres au cloud

Pourquoi le sharding ?

Facebook en 2004 -> un serveur

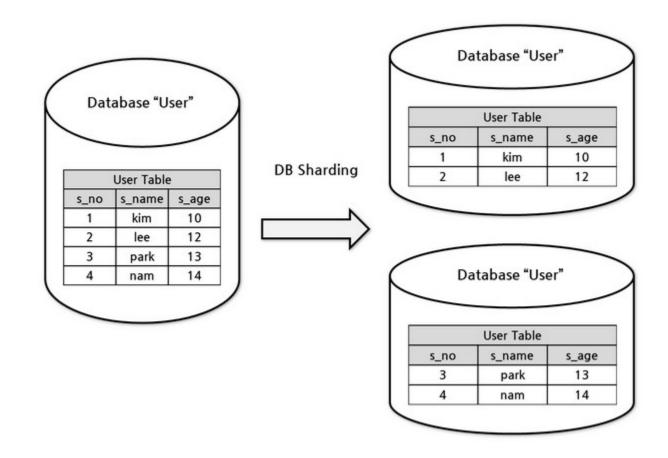
Facebook en 2008 -> 5000 pages par seconde

Les metadata des photos (identifiants, localisation) représentent à elles seules des teras de données!

Blocage au niveau des I/O, problèmes de scale-up

Impact :

- La connexion à la base nécessite une étape de plus : identifier la bonne base
- Les données ne sont plus normalisées
- Les indisponibilités sont limitées
- C'est le soft qui s'occupe de la cohérence



Impact :

- La connexion à la base nécessite une étape de plus : identifier la bonne base
- Les données ne sont plus normalisées
- Les indisponibilités sont limitées
- C'est le soft qui s'occupe de la cohérence

L'architecture des applications dans le cloud

- Cloud bursting :
 - Problématiques :
 - Quelle architecture :
 - Clusters indépendants ?
 - Source de donnée unique / multiple ?
 - Latences de l'environnement de distant ?
 - Où stocker les données (en local, dans le cloud, les deux?)

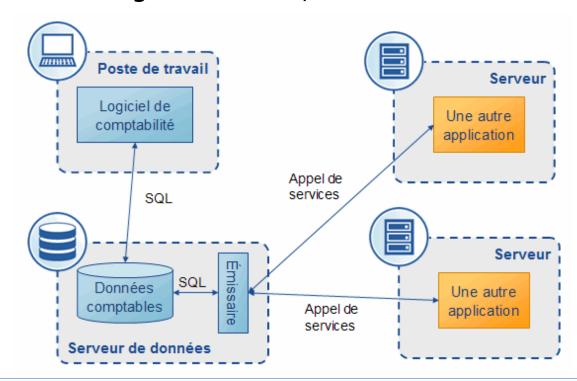
Haute disponibilité applicative dans le cloud

- Même si les fournisseurs de cloud font le maximum, les pannes ne sont pas exclues, exemple avec Amazon :
 - 20 avril 2011 : panne sur EBS, blocage des I/O
 - 29 Juin 2012 : Tornade + problèmes de reprise d'activité
 - 22 octobre 2012 : problème dormant sur EBS
 - Noël 2012 : problème sur l'Elastic Load Balancing
- Il faut mettre en place de la redondance en place et de la distribution de ressources
 - -> On se repose uniquement sur du soft car on ne plus utiliser les technologies de stockages ou de virtualisation proposées dans le monde IT classique

Haute disponibilité applicative dans le cloud

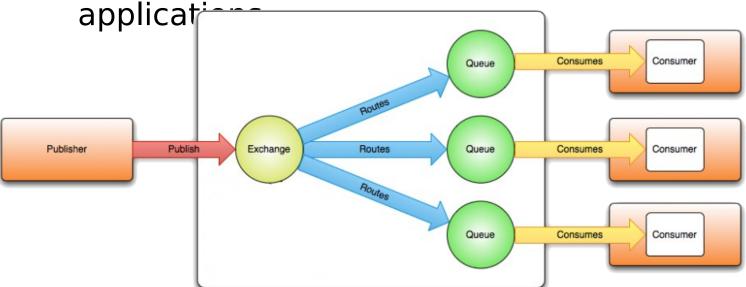
Le SOA

 Forme d'architecture de médiation qui est modèle d'interaction applicative qui met en œuvre des services (composants logiciels) « Wikipedia »



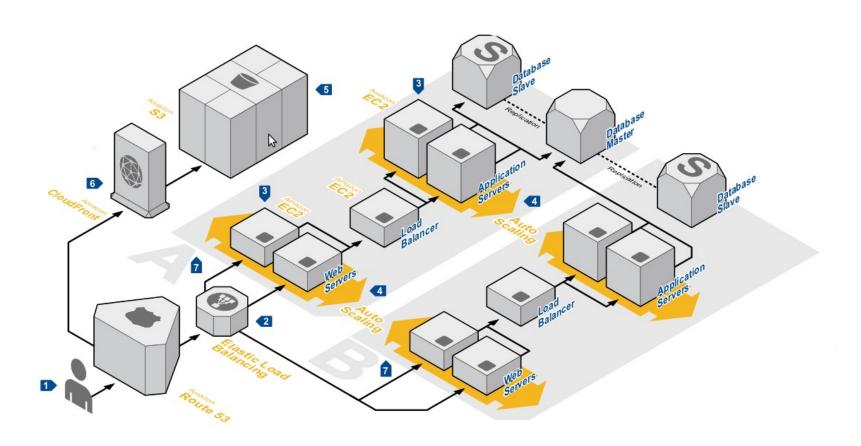
- Interprocess communication
 - Advanced Message Queuing Protocol (AMQP)

Protocole de messagerie destinée au applications

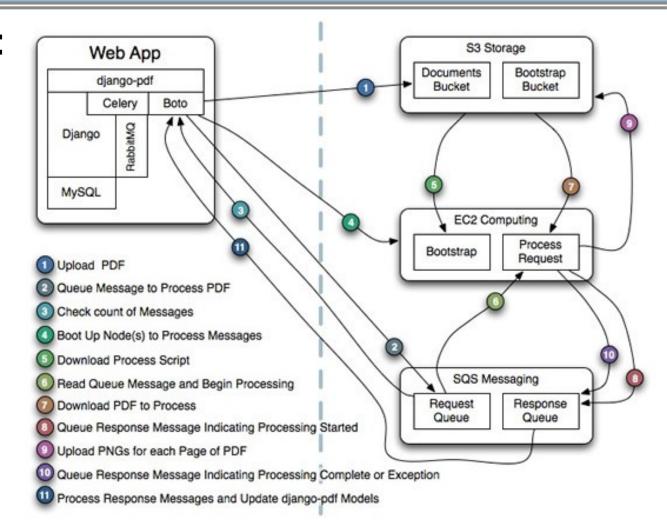


Haute disponibilité applicative dans le cloud

Exemple :



Exemple :



- Tests de résilience
 - Chaos Gorilla Army (Netflix) pour simuler des pannes
- Tests de conformité
 - Chaos Monkey (Netflix) pour vérifier les configurations, les gaspillages, la sécurité

Exploiter dans le cloud

- Utiliser les outils et méthodes issus du mouvement « devops »
 - Build one, deploy many
 - Toutes les configurations doivent être sous contrôle de version
 - Maximum de standardisation et d'automatisation

Mesurer

Exploiter dans le cloud

- Méthodologies :
 - ITIL
 - IT Capability Maturity Framework (IT-CMF)
 - Permet d'évaluer si l'organisation est prête à passer dans le cloud

Le cloud en pratique

- 1) Choisir ce que l'on souhaite mettre dans le cloud
 - Applications éligibles, simulation des coûts, ...
- 2) Choisir le type
 - Public / Privé / Hybride
- 3) Choisir son fournisseur
 - Réputation, agréments, ...
- 4) Définir le contrat
 - SLA, scénarios de ré-internalisation, ..

Le futur du cloud

- De moins en moins cher et plus sécurisé
- Des mega datacenter répartis dans le monde
- Automatisation croissante des actions
- Standardisation des API
- PaaS et FaaS domineront

48

Le futur du cloud

- Un cloud mondial ?
- Retour en arrière suite à un crash de grande ampleur ?
- Disparition de certains métiers liés à l'infrastructure ?
- Naissance de skynet ?

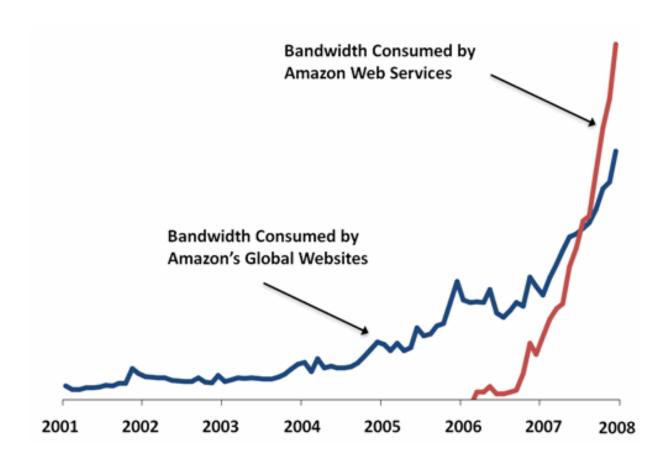
49

Amazon Web Services (AWS)

- Présentation
- Les Amazon Machine Images
- Les Instances
- Réseaux et sécurité
- Stockage
- Ressources et tags

- Naissance en 2002
- AWS se composent de plusieurs services :
 - Elastic Compute Cloud (EC2)
 - Fourniture de VM
 - Elastic Block Store (EBS)
 - Stockage bloc persistant pour les instances EC2
 - Simple Storage Service (S3)
 - Stockage basé sur les services Web
 - Simple Queue Service (SQS)
 - AWS Management Console (WAS Console)

Amazon Web Services



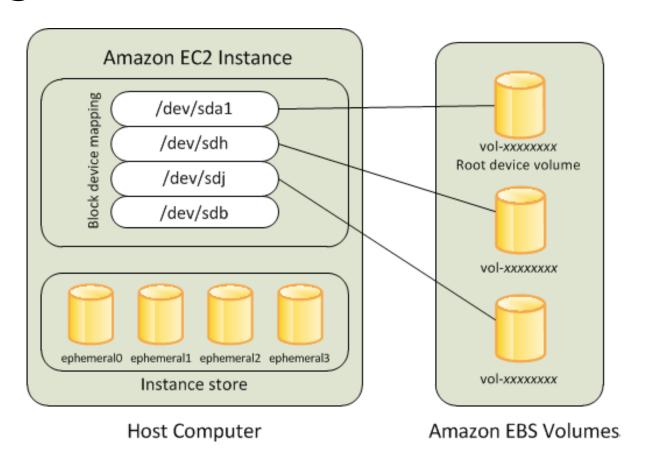
Les Amazon Machine Images

- Aussi nommées AMI
 - Templates que l'on peut créer ou importer (gratuit ou payant)

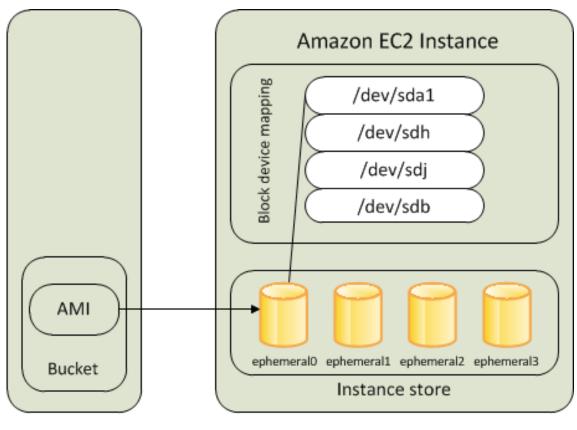
- 2 backend de stockage :
 - EBS
 - Instance store-backed

Les Amazon Machine Images

EBS

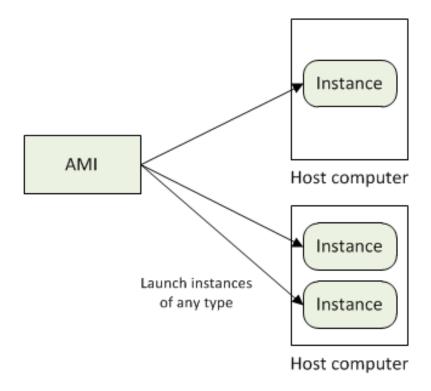


Instance store-backed



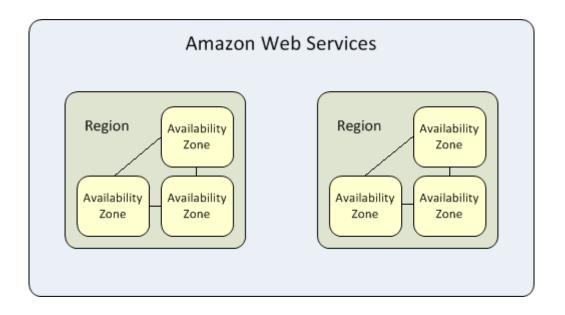
Amazon S3

 Une copie d'une AMI qui fonctionne dans le cloud



- Plusieurs types d'instances :
 - Cluster Compute, Cluster GPU, High CPU, ...
 - Le type d'instance définit le type de configuration
 - CPU, RAM, disque, réseau, architecture, ...
- Les instances fonctionnent à partir d'une
 « Availability Zone »

 Une Région est composée de plusieurs « Availability Zones »



| Region |
|--------------------------------------|
| US East (Northern Virginia) Region |
| US West (Oregon) Region |
| US West (Northern California) Region |
| EU (Ireland) Region |
| Asia Pacific (Singapore) Region |
| Asia Pacific (Sydney) Region |
| Asia Pacific (Tokyo) Region |
| South America (Sao Paulo) Region |

- Les « Availability Zones » d'une même région sont isolées les unes des autres et permettent de limiter les pannes.
- Elles sont interconnectées par des réseaux à faible latence
- Le transfert de données est gratuit entre VM dans la même AZ
- Le transfert de données entre AZ différentes est payant
- Le transfert de données entre région est payant

- Accès à la console :
 - Via le web (client Java)
 - Via SSH (Linux)
 - Via RDP (Windows)
 - Via http après installation de logiciels
 - Pour connaitre l'IP :
 - console EC2
 - ec2-describe-instances (API Tools)

- Autoscaling
 - Permet d'arrêter ou démarrer des instances en fonction de certains critères (CPU,RAM)
 - On crée :
 - une politique de lancement
 - un groupe et une politique d'ajustement
 - Coûts additionnels (CloudWatch)

Coûts des instances (2013):

| Instances standards à la demande | | | |
|--|--------------------|--------------------|--|
| Small (par défaut) | \$0,060 de l'heure | \$0,091 de l'heure | |
| Medium | \$0,120 de l'heure | \$0,182 de l'heure | |
| Large | \$0,240 de l'heure | \$0,364 de l'heure | |
| Extra Large | \$0,480 de l'heure | \$0,728 de l'heure | |
| Instances à la demande standard de deuxième génération | | | |
| Extra Large | \$0,500 de l'heure | \$0,780 de l'heure | |
| Double Extra Large | \$1,000 de l'heure | \$1,560 de l'heure | |
| Instances micro à la demande | | | |
| Micro | \$0,020 de l'heure | \$0,020 de l'heure | |
| Instances à la demande à mémoire élevée | | | |
| Extra Large | \$0,410 de l'heure | \$0,510 de l'heure | |
| Double Extra Large | \$0,820 de l'heure | \$1,020 de l'heure | |
| Quadruple Extra Large | \$1,640 de l'heure | \$2,040 de l'heure | |

Coûts réseaux

- IP:
- \$0,00 pour une adresse IP élastique associée à une instance active
- \$0,005 par adresse IP élastique supplémentaire associée à une instance active ; base horaire au prorata
- \$0,005 par adresse IP élastique non associée à une instance active ; base horaire au prorata
- \$0,00 par remappage d'adresse IP élastique 100 premiers remappages/mois
- \$0,10 par remappage d'adresse IP élastique remappages supplémentaires au-delà de 100/mois

Load Ralancing:

- \$0,025 par Elastic Load Balancer-heure (ou heure partielle)
- \$0,008 par Go de données traitées par un Elastic Load Balancer

Coûts de stockage :

Volumes Standard Amazon EBS

- \$0,10 par Go-mois de stockage alloué
- \$0,10 par 1 million de demande E/S

Volumes IOPS dimensionnés Amazon EBS

- \$0,125 par Go-mois de stockage alloué
- \$0,10 par IOPS dimensionné et par mois

Instantanés Amazon EBS vers Amazon S3

\$0,095 par Go-mois de données stockées

- Les security groups
 - Agissent comme des firewalls
 - Sont composés de règles
 - Gérés par les hyperviseurs d'Amazon
 - Ils sont propres à chaque Région
 - Par défaut
 - tout le trafic entrant est bloqué et tout le trafic sortant autorisé vers l'extérieur
 - tout types de trafic est autorisé entre les VMs d'un même sécurity group
 - Sur EC2 classique, il faut un reboot de l'instance en cas de modification de règles

Réseaux et sécurité

- Contrôles d'accès :
 - Identity and Access Management (IAM)
 - Gestion de groupes et d'utilisateurs
 - Gestion des ressources par utilisateur/groupe
 - Upload de certificat pour l'utilisation de l'API
 - Login / mot de passe pour les interfaces de management

Réseaux et sécurité

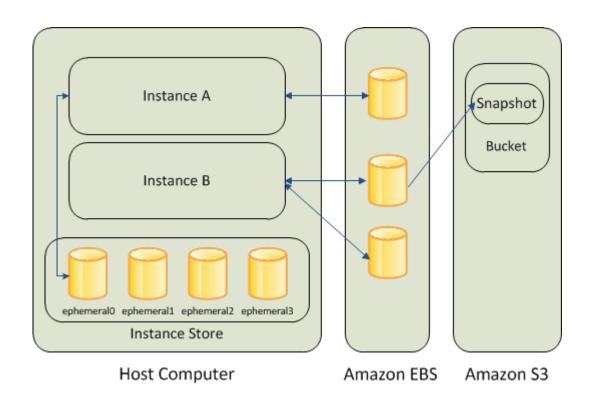
Adressage IP :

- Attribution d'IP privée + enregistrement DNS + ip publique
- Communication avec l'extérieur via du NAT

Elactic IP

 Adresse IP que l'on peut attribuer à n'importe quelle instance

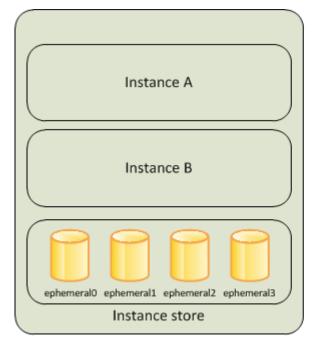
Vue générale :



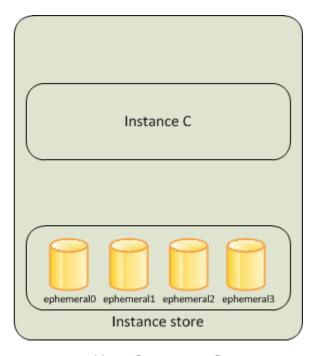
- Elastic Block Sotrage
 - Périphérique de type bloc
 - Utilisable avec les instances
 - Persiste dans le temps
 - Différents niveaux de performances
 - Peut être déplacé entre deux VM dans la même AZ
 - Supporte les snapshots
 - On déplacer les volumes entre AZ via le service S3

Instance storage :

Volatile



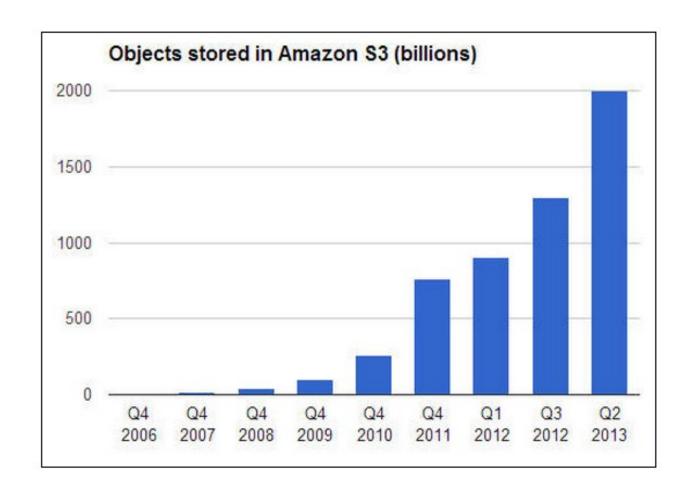
Host Computer 1



Host Computer 2

- Amazon S3:
 - Entrepôt de données
 - Faible coût
 - Notion de « bucket »
 - Composé de metadata et de data

- Exemple :
 - http://myawbucket.s3.amazonaws.com/photos/ monjardin.jpg



Ressources et tags

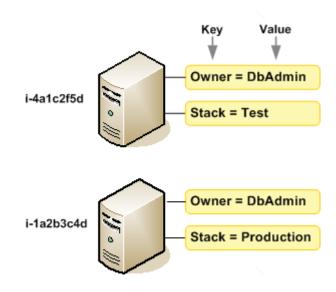
Les ressources :

| Resource | Туре | Description |
|----------------------------------|-----------------------|---|
| AWS Account | Global | You use the same AWS account in all Regions. |
| DevPay Product Codes | Global | You use the same DevPay product codes throughout all Regions. |
| Amazon EC2 System Identifiers | Regional | Includes the AMI ID, Instance ID, EBS Volume ID, EBS Snapshot ID, and so on. |
| Instances | Availability Zone | Instances are tied to Availability Zones. However, the instance ID is tied to the Region. |
| AMIs | Regional | AMIs are tied to the Region where its files are located within Amazon S3. |
| Security Groups | Regional | Security groups are not copied across Regions. Instances within the Region cannot communicate with instances outside the Region using group-based firewall rules. Traffic from instances in another Region is seen as WAN bandwidth. |
| SSH Key Pairs | Regional or Global | The SSH key pairs that you create with ec2-add-keypair, CreateKeyPair, or in the AWS Management Console work only in the Region where you create them. However, you can optionally create an RSA key pair with a third-party tool and upload the public key to AWS. That key pair works in all Regions. For more information, go to ec2-import-keypair in the Amazon Elastic Compute Cloud Command Line Reference or ImportKeyPair in the Amazon Elastic Compute Cloud API Reference. |
| User-Supplied Identifiers | Regional | Includes security group names, SSH key pair names, and so on. Although you can create the same names in multiple Regions, they have no relationship to each other. |
| Elastic IP Addresses | Regional | Elastic IP addresses are tied to a Region and cannot be mapped across Regions. |
| EBS Volumes | Availability Zone | An Amazon EBS volume must be located within the same Availability Zone as the instance to which it attaches. |
| EBS Snapshots | Regional | Snapshots are tied to Regions and can only be used for volumes within the same Region. |

Ressources et tags

Les tags :

- Permet de faciliter l'administration
- Maximum de tags par ressources : 10
- Taille maximum : 256 caractères



Introduction à Apache CloudStack

- Présentation
- Historique
- Architecture et composants
- Terminologie (régions, zones, pod, ..)
- Mise en place d'une plateforme minimale

- Logiciel permettant de créer et gérer une infrastructure de cloud computing de type laaS (CMP)
- Il gère l'orchestration des ressources :
 - Virtualisation
 - Stockage
 - Réseau
- Il propose sa propre API mais supporte aussi celle d'Amazon (AWS) et de l'Open Grid Forum (OCCI)

Historique

- Naissance un peu avant 2010 sous le nom de VMOps (cloud.com)
- Racheté par Citrix en Juillet 2011 -> version 3.0
- Donné à la fondation Apache en Avril 2012 -> version 4.0
- Version actuelle : 4.3 (Mars 2014)

Architecture et composants

- Pour faire fonctionner une plateforme cloudstack, il nous faut :
 - un serveur ou de(s) serveur(s) de management (Controller)
 - des hyperviseurs (Kvm, Xen, vSphere, HyperV)
 - du stockage (primaire et secondaire)
 - du réseau (basique ou avancé)

Terminologie

Region

 Equivalent à une région géographique (1 à n zones)

Zone

- Equivalent à un datacenter (1 à n pod)
- Attribution du stockage secondaire
- Type de réseau

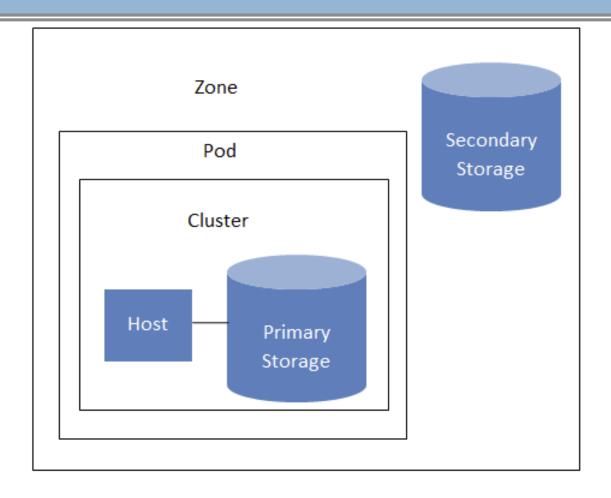
Pod

Equivalent à un rack (1 à n clusters)

Terminologie

- Cluster
 - Groupe d'hôtes (1 à n)
 - Ils partagent un stockage primaire
- Host
 - •

- Instance
 - Une VM



Nested organization of a zone

Le stockage primaire

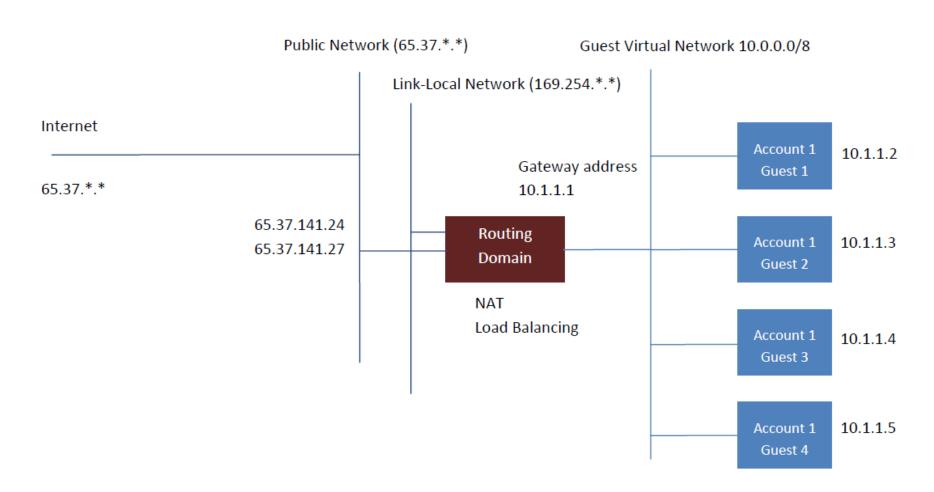
- Stocke les VM
- Même problématique que pour le stockage des VM en virtualisation (performance, disponibilité, ..)
- Protocoles supportés :
 - SCSI (DAS), FC, iSCSI, NFS
- Systèmes de fichiers partagés supportés :
 - VMFS (VMware), CLVM, Ceph, RDB, GFS2, OCFS2, NFS
- En cas de stockage local (DAS) :
 - Ext4, xfs, ...

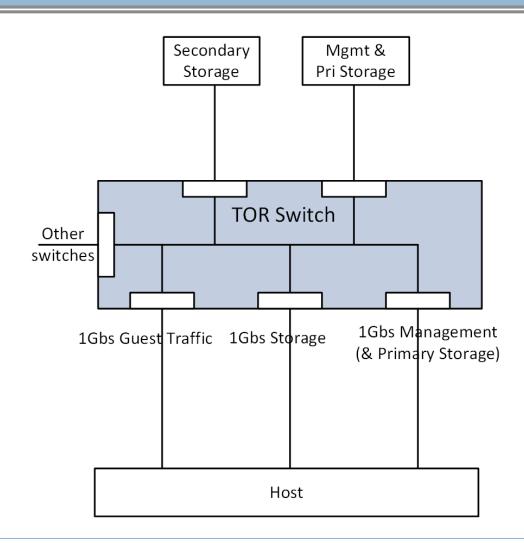
Le stockage secondaire

- Stocke :
 - Les Templates
 - Les ISO
 - Les snapshots (en fonction du stockage primaire)
 - Les volumes importés
 - Les volumes des VM lors des migration

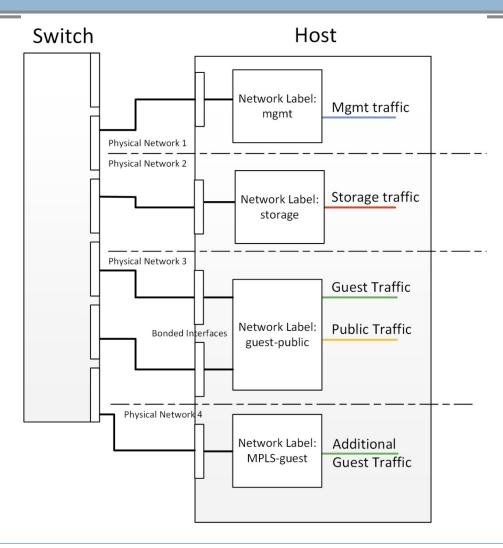
- Supporte NFS ou S3
- Ne nécessite pas un stockage rapide

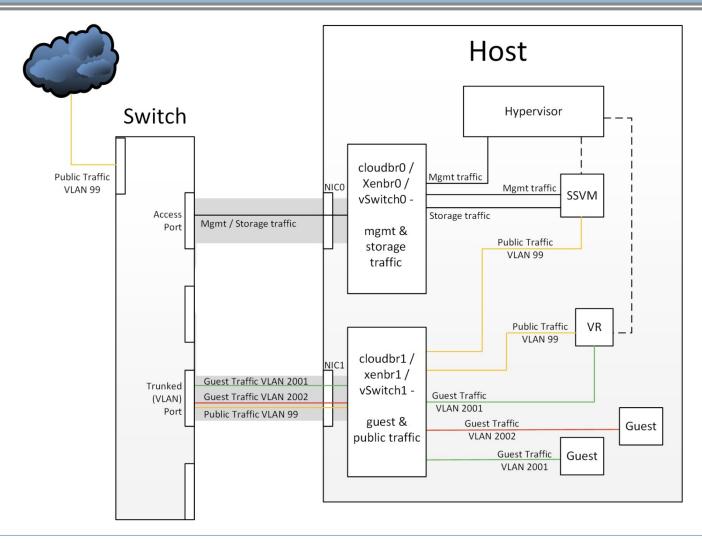
- Mode basique :
 - Trafic à « plat »
 - Utilisation de security groups (isolation L3)
 - Configuration assez simple
 - Support Elastic IP (EIP) / Elastic LoadBalancer (ELB) via des netscaler





- Mode avancé :
 - VLAN ou SDN pour l'isolation (L2)
 - Pas de security groups
 - Support des Virtual Private Cloud
 - Support des VPN
 - Routage inter-vlan
 - Configuration plus complexe





VM système

- Elles sont au nombre de trois :
 - Virtual router (VR)

Secondary storage VM (SSVM)

Console Proxy VM (CPVM)

Mise en place d'une plateforme

- 1) Conception de l'architecture
- 2) Préparation de l'environnement :
 - Installation du contrôleur (VM)
 - Installation des hyperviseurs (Mise en rack, configuration)
 - Préparation du stockage (création des volumes, ACL)
 - Préparation du réseau (Vlan, routage, QoS, ..)

Mise en place d'une plateforme

- 3) Configuration CloudStack
 - Région -> Zone (premier pod)
 - Cluster
 - Stockage primaire -> stockage secondaire
 - Offre(s) de service
 - Offre réseau
- 4) Pour aller plus loin
 - Ajouter des utilisateurs
 - Créer des templates
 - Mettre en place des VPN, du loadbalancing